Vypracované otázky na skúšku z VŠA 1

Otázka 1. Uveďte definíciu náhodného vektora a jeho distribučnej funkcie. Skonštruujte príklad funkcie z $\mathbb{R}^2$ do $\mathbb{R}$, ktorá je distribučnou funkciou nejakého náhodného vektora. Definujte kovariančnú maticu náhodného vektora a uveďte jej základné vlastnosti. Definujte kovarianciu dvoch náhodných vektorov.

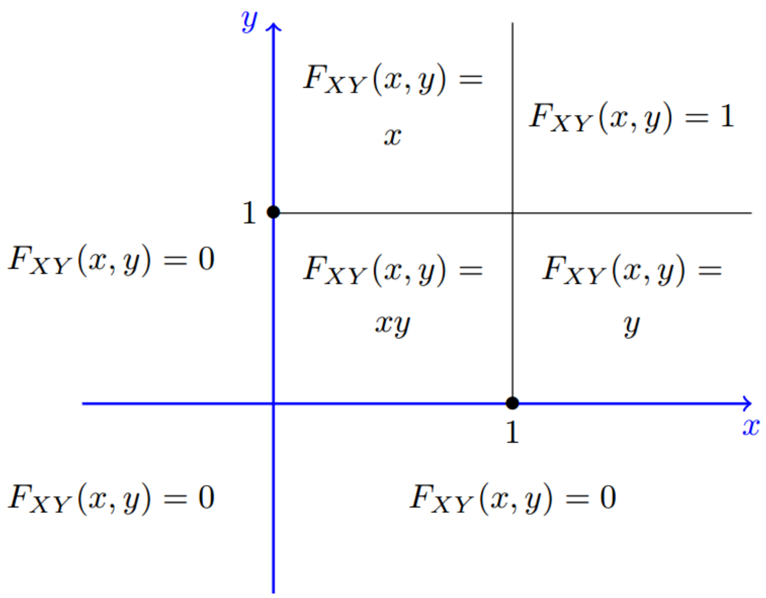
Nech (Ω, S, P) je pravdepodobnostný priestor. Náhodný vektor $X = (X_1, \dots, X_n)^T$ nazývame také zobrazenie $X : \Omega \to \mathbb{R}^n$, že

$$\forall x_i \subset (x_1, \dots, x_n)^T \in \mathbb{R}^n : \bigcap_{i=1}^{n} (\omega : X_i(\omega) < x_i) \in S$$

Distribučnú funkciu X nazývame

$$F : \mathbb{R}^n \to \mathbb{R} : \; F(X_1, \dots, X_n) = P\left(\bigcap_{i=1}^{n} (\omega : X_i(\omega) < x_i)\right) = P(X < x)$$

Príklad funkcie z $\mathbb{R}^2$ do $\mathbb{R}$, ktorá je distribučnou funkciou nejakého náhodného vektora:
Nech $X, Y \sim R(0,1)$, potom distribučná funkcia $F_{XY}(x,y)$ náhodného vektora $(X,Y)^T$ je



Kovariančná matica náhodného vektora je

$$\begin{pmatrix} Var(X_1) & \dots & Cov(X_1, X_n) \\ \vdots & \vdots & \vdots \\ Cov(X_n, X_1) & \dots & Var(X_n) \end{pmatrix}$$

Vlastnosti kovariančnej matice sú

$$\begin{aligned} Var(X) &\geq 0 \\ Var(X) &= E(XX^T) - E(X)E^T(X) \\ Var(a^T X) &= a^T Var(X) a \\ Var(AX + b) &= A Var(x) A^T \\ Cov(X + Y, Z) &= Cov(X, Z) + Cov(Y, Z) \\ Cov(AX, BY) &= A Cov(X, Y) B^T \\ Var(X + Y) &= Var(X) + Var(Y) + Cov(X, Y) + Cov(Y, X) \end{aligned}$$

Kovariancia dvoch náhodných vektorov je

$$Cov(X,Y) = E[(X - E(X))(Y - E(Y))]$$

Otázka 2. Nech X_1, X_2 sú náhodné vektory a nech existuje $E(X_1)$. Zadefinujte podmienenú strednú hodnotu a podmienenú varianciu náhodného vektora X_1 za podmienky, že $X_2 = x_2$. Nech X_1, X_2 sú náhodné vektory a nech existuje $E(X_1)$. Popíšte aproximáciu X_1 funkciou X_2 pomocou podmienenej strednej hodnoty $E(X_1|X_2)$.
Podmienená stredná hodnota náhodného vektora je

$$E(X_1|X_2 = x_2) = \int_{\mathbb{R}} x_1 f(X_1|X_2 = x_2)\, dx_1,$$

kde

$$f(X_1|X_2 = x_2) = \frac{f(x_1, x_2)}{f(x_2)}$$

Podmienená variancia náhodného vektora je

$$Var(X_1|X_2 = x_2) = E(X_1 X_1^T|X_2 = x_2) - E(X_1|X_2 = x_2)E^T(X_1|X_2 = x_2)$$

Pre každú merateľnú funkciu $h(X_2) = E(X_1|X_2)$ platí

$$E[X_1 - h(X_2)]^2 \geq E[X_1 - E(X_1|X_2)]^2$$

Rovnosť nastáva práve vtedy, keď $h(X_2) = E(X_1|X_2)$ s pravdepodobnosťou 1. To je všeobecná aproximácia X_1 funkciou X_2 pomocou podmienenej strednej hodnoty $E(X_1|X_2)$.

Otázka 3. Uvažujme náhodné premenné X, Y s distribučnými funkciami F, G a združenou distribučnou funkciou H. Definujte kopulu týchto náhodných premenných a popíšte vzťah medzi kopulou C a distribučnými funkciami F, G, H.
$C : \langle 0,1\rangle^2 \to \mathbb{R}$ nazývame 2-rozmernou kopulou, ak platia nasledovné vlastnosti:

1.
$$C(u,0) = 0$$
$$C(0,v) = 0$$

2.
$$C(u,1) = u$$
$$C(1,v) = v$$

3.
$$\text{objem kopule } V_C([u,v]) \geq 0$$

Nech H je 2-rozmerná združená distribučná funkcia s marginálnymi distribučnými funkciami F, G a kvantilovými funckiami F^{-1}, G^{-1}. Potom existuje 2-rozmerná kopula C taká, že:

$$H(x,y) = C(F(x), G(y))$$

a

$$C(u,v) = H(F^{-1}(u), G^{-1}(v))$$

Otázka 4. Nech C je n-rozmerná kopula. Popíšte jej ohraničenia (dolné a horné) a stručne popíšte ich odvodenie pre $n = 2$.
Dolné ohraničenie

$$W^{(n)}(u_1, \ldots, u_n) = \max\{u_1 + \cdots + u_n - n + 1, 0\}$$

Odvodenie pre $n = 2$: Nech $U = 1 - V$

$$W(u,v) = P[U < u, V < v] = P[U < u, 1 - U < v] = P[U < u, U > 1 - v] = \begin{cases} u + v - 1 & \text{ak } 1 - v < u \\ 0 & \text{ak inak} \end{cases}$$

Horné ohraničenie

$$M^{(n)}(u_1, \ldots, u_n) = \min\{u_1, \ldots, u_n\}$$

Odvodenie pre $n = 2$: Nech $U = V$

$$M(u,v) = P[U < u, V < v] = \min\{u, v\}$$

Otázka 5. Nech p-rozmerný náhodný vektor X má regulárne viacrozmerné normálne rozdelenie $N_p(\mu, \Sigma)$. Akú má X hustotu, strednú hodnotu a kovariančnú maticu? Aké rozdelenie majú zložky vektora X? Aké rozdelenie má vektor $AX + b$, kde $b \in \mathbb{R}^k$ a A je matica typu $k \times p$?
Hustota viacrozmerného normálneho rozdelenia X je

$$f(x) = \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu)\right),$$

Stredná hodnota viacrozmerného normálneho rozdelenia X je μ.
Kovariančná matica viacrozmerného normálneho rozdelenia X je Σ.
Zložky vektora X majú tiež normálne rozdelenie $X_i \sim N_1(\mu_i, \Sigma_{ii})$.
Vektor $AX + b$ má rozdelenie $N_q(A\mu + b, A\Sigma A^T)$.

Otázka 6. Nech $X_1, \ldots, X_n$ je náhodný výber z p-rozmerného rozdelenia so strednou hodnotou μ a kovariančnou maticou Σ. Definujte základné výberové charaktristiky a uveďte ich strednú hodnotu.
Výberový priemer je

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Stredná hodnota výberového priemeru je

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \frac{1}{n} n\mu = \mu$$

Výberová kovariančná matica je

$$S = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})^T$$

Stredná hodnota výberovej kovariančnej matice je

$$
\begin{aligned}
E(S) &= E\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})^T\right) \\
&= E\left(\frac{1}{n}\sum_{i=1}^{n}X_iX_i^T - \bar{X}\bar{X}^T\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}E(X_iX_i^T) - E(\bar{X}\bar{X}^T) \\
&= (\Sigma + \mu\mu^T) - \left(\frac{1}{n}\Sigma + \mu\mu^T\right) \\
&= \frac{n-1}{n}\Sigma
\end{aligned}
$$

Otázka 7. Nech $X \sim N_p(0, \Sigma)$, A je matica typu $p \times p$ a $A\Sigma$ je idempotentná. Uveďte, aké rozdelenie má kvadratická forma X^TAX a vypočítajte jej strednú hodnotu.
Ak $A\Sigma$ je idempotentná, potom kvadratická forma X^TAX má Chí-kvadrát rozdelenie χ_r^2, kde $tr(A\Sigma) = r$.

$$
\begin{aligned}
E(X^TAX) &= E(tr(X^TAX)) \\
&= E(tr(AXX^T)) \\
&= tr(AE(XX^T)) \\
&= tr(A(\Sigma + \mu\mu^T)) \\
&= tr(A\Sigma) + tr(A\mu\mu^T) \\
&= tr(A\Sigma) + tr(\mu^TA\mu) \\
&= tr(A\Sigma) + \mu^TA\mu
\end{aligned}
$$

Keďže $\mu = 0$, tak $E(X^TAX) = tr(A\Sigma)$.

Otázka 8. Definujte Wishartovo rozdelenie. Ukážte, že ak $M \sim W_p(\Sigma, n)$, potom $E(M) = n\Sigma$.
Nech $X_1, \dots, X_n$ je náhodný výber z $N_p(0, \Sigma)$. Nech $X = (X_1, \dots, X_n)^T$. Potom rozdelenie matice $M = X^TX$ nazývame Wishartovým rozdelením s n stupňami voľnosti.

$$
E(M) = E(X^TX) = E\left(\sum_{i=1}^{n}X_iX_i^T\right) = nE(X_iX_i^T) = n[D(X_i) + E(X_i)E(X_i^T)] = n\Sigma
$$

Otázka 9. Nech $X_1, \dots, X_n$ je náhodný výber z $N_p(0, \Sigma)$. Definujte výberovú kovariančnú maticu a uveďte jej rozdelenie. Nech $X_1, \dots, X_n$ je náhodný výber z $N_p(\mu, \Sigma)$ a $Y_i = AX_i$, $i = 1, \dots, n$. Vyjadrite výberový priemer a výberovú kovariančnú maticu Y pomocou výberového priemeru a výberovej kovariančnej matice X.
Výberová kovariančná matica X je

$$
S_X = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})^T
$$

Rozdelenie výberovej kovariančnej matice je

$$
S \sim \frac{1}{n}W_p(\Sigma, n-1)
$$

Výberový priemer Y je

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n}\sum_{i=1}^{n} AX_i = \frac{A}{n}\sum_{i=1}^{n} X_i = A\bar{X}$$

Výberová kovariančná matica Y je

$$S_Y = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})(Y_i - \bar{Y})^T = \frac{1}{n}\sum_{i=1}^{n}(AX_i - A\bar{X})(AX_i - A\bar{X})^T = \frac{A^2}{n}\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})^T = A^2 S_X$$

Otázka 10. Zadefinujte Fisherovu informačnú maticu a uveďte dolné ohraničenie pre varianciu odhadu $\hat{\theta} = t(X)$ parametra θ.
Fisherová informačná matica je kovariančná matica skórovej funkcie.

$$F_n = Var\left[\frac{\partial l(X,\theta)}{\partial\theta}\right] = -E\left[\frac{\partial^2 l(X,\theta)}{\partial\theta\partial\theta^T}\right]$$

kde $l(X,\theta) = \ln L(X,\theta) = \ln(f(x_{11},\ldots,x_{1n}).\ldots.f(x_{n1},\ldots,x_{nn}))$.
Ak $\hat{\theta} = t(X)$ je ľubovoľný nevychýlený odhad θ, potom za podmienok regularity platí

$$D(t(X)) \geq F_n^{-1}$$

Otázka 11. Nech $X = (X_1,\ldots,X_n)^T$ je náhodný výber z rozdelenia závislého od parametra $\theta \in \mathbb{R}^k$. Zadefinujte test pomerom vierohodnosti pre $H_0: \theta \in \Theta_0$ oproti $H_1: \theta \in \Theta_1$ na hladine významnosti α. Uveďte tvar kritickej oblasti pre tento prípad, tvar testovacej štatistiky a jej asymptotické rozdelenie.
Test pomerom vierohodnosti má testovaciu štatistiku

$$\lambda(x) = \frac{L_0^*}{L_1^*},\ L_i^* = \max_{\theta\in\Theta_i} L(X,\theta)$$

Resp.

$$-2\ln(\lambda(x)) = 2(l_1^* - l_0^*),$$

kde $-2\ln(\lambda(x)) \sim \chi^2_{q-r}$.
Kritická oblasť likelihood ratio testu na hladine významnosti α je

$$W = \{x : \lambda(x) < c\}$$

Resp.

$$W = \{x : -2\ln(\lambda(x)) > k\}$$

Otázka 12. Uvažujme dva nezávislé náhodné výbery rozsahu n z dvojrozmerného normálneho rozdelenia $N_2(\mu,\Sigma)$. Uveďte test pomerom vierohodnosti pre hypotézu $H_0: \mu_1 = \mu_2$, kde $\mu = (\mu_1,\mu_2)^T$.
1. Vypočítame $\bar{X}_1, \bar{X}_2, S_1, S_2$.
2. Vieme, že:

$$\bar{X}_1 - \bar{X}_2 \sim N_2(\mu_1 - \mu_2, \frac{1}{n_1}\Sigma + \frac{1}{n_2}\Sigma)$$

3. Výberová variančná matica je:

$$n_1 S_1 + n_2 S_2 \quad \sim \quad W_2(\Sigma, n_1 + n_2 - 2)$$

Označme si

$$S \quad = \quad \frac{1}{n_1 + n_2}(n_1 S_1 + n_2 S_2)$$

kde S je nezávislé od $(\bar{X}_1 - \bar{X}_2)$.
4. Testovacia štatistika je

$$\frac{n_1 n_2 (n_1 + n_2 - 2)}{(n_1 + n_2)^2}(\bar{X}_1 - \bar{X}_2)^T S^{-1} (\bar{X}_1 - \bar{X}_2) \quad \sim \quad T^2(2, n_1 + n_2 - 2)$$

alebo

$$(\bar{X}_1 - \bar{X}_2)^T S^{-1} (\bar{X}_1 - \bar{X}_2) \quad \sim \quad \frac{2(n_1 + n_2)^2}{n_1 n_2 (n_1 + n_2 - 2 - 1)} F(2, n_1 + n_2 - 2 - 1)$$

Otázka 13. Uvažujme dva náhodné výbery $X_1, \ldots, X_n$ z $N_p(\mu_1, \Sigma)$ a $Y_1, \ldots, Y_n$ z $N_p(\mu_2, \Sigma)$, medzi ktorými existuje prirodzené párovanie. Odvoďte test pomerom vierohodnosti pre hypotézu $H_0: \ \mu_1 = \mu_2$.

$$\begin{pmatrix} X \\ Y \end{pmatrix} \quad \sim \quad N_{2p}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right)$$

Nech $d_i = X_i - Y_i$. Potom

$$\mu_d = E(X - Y) = \mu_1 - \mu_2$$

Testovaciu hypotézu môžeme upraviť na $H_0: \ \mu_d = 0$. Ďalej

$$\bar{d} \quad = \quad \frac{1}{n}\sum_{i=1}^{n} d_i$$
$$S_d \quad = \quad \frac{1}{n}\sum_{i=1}^{n}(d_i - \bar{d})(d_i - \bar{d})^T$$

Testovacia štatistika je

$$(n-1) d^T S^{-1} d \quad \sim \quad T^2(p, n-1)$$

Otázka 14. Uvažujme lineárny regresný model tvaru $Y_{n\times p} = X_{n\times q} B_{q\times p} + \varepsilon_{n\times p}$. Popíšte spôsob testovania lineárnej hypotézy o neznámych parametroch $H_0: \ LB = 0$, kde L je matica typu $r \times p$. Testovaním zisťujeme, buď či niektoré faktory X_i majú vplyv na pozorovania, t. j. $H_0: \ \beta_i = 0$, alebo testujeme rovnosť vplyvu faktorov na pozorovania, t. j. $H_0: \ \beta_i = \beta_j$.

$$H_{p\times p} \quad = \quad \hat{B}^T L^T [L(X^T X)^{-1} L^T]^{-1} L \hat{B}$$
$$E_{p\times p} \quad = \quad \varepsilon^{\hat{T}} \varepsilon = Y^T Y - \hat{B}^T (X^T X)^{-1} \hat{B}$$

Usporiadame vlastné hodnoty matice $E^{-1}H$: $\lambda_1 \geq \cdots \geq \lambda_p \geq 0$ a testujeme pomocou niektorej testovacej štatistiky.

Otázka 15. Uveďte najčastejšie používané testovacie štatistiky pre testy hypotéz vo viacrozmernom regresnom modeli.
Pillaiovo stopové kritérium:

$$P = tr[(E+H)^{-1}E] = \sum_{i=1}^{p} \frac{\lambda_i}{1+\lambda_i}$$

Wilksovo Λ-kritérium:

$$W = det|(E+H)^{-1}E| = \prod_{i=1}^{p} (1+\lambda_i)^{-1}$$

Hotelling-Lawleyvo stopové kritérium:

$$HL = tr(E^{-1}H) = \sum_{i=1}^{p} \lambda_i$$

Royvo kritérium najväčšej vlastnej hodnoty:

$$R = \max_{\lambda}(E^{-1}H) = \lambda_1$$

Otázka 16. Nech Y_{ij} je j-te pozorovanie, $j = 1, \dots, n_i$ v i-tom základnom súbore $N(\mu_i, \sigma^2)$, $i = 1, \dots, I$. Testujeme hypotézu $H_0: \ \mu_1 = \dots = \mu_I (= \mu)$. Zostavte zodpovedajúci model jednofaktorovej analýzy rozptylu pre tento prípad a uveďte postup testovania hypotéz.
Predpoklady modelu sú:

- nezávislosť
- normalita
- homoskedasticita

Platí: $Y_{ij} = \mu_i + \varepsilon_i$, kde ε_i je náhodná chyba z $N(0, \sigma^2)$. Ďalej nech

$$\begin{aligned} Y_{i.} &= Y_{i1} + \dots + Y_{in_i} \\ Y_{..} &= Y_{1.} + \dots + Y_{I.} \\ SS_H &= \sum_i n_i \left(\bar{Y}_{i.} - \bar{Y}_{..}\right)^2 \\ SS_E &= \sum_i \sum_j \left(Y_{ij} - \bar{Y}_{i.}\right)^2 \end{aligned}$$

1. $H_0: \ \mu_1 = \dots = \mu_I (= \mu)$
Testovacia štatistika je

$$F = \frac{\frac{SS_H}{I-1}}{\frac{SS_E}{n-I}} \sim F\left(I-1, n-I\right)$$

2. Zistiť, ktoré rozdiely $\mu_i - \mu_j$ sú rôzne od nuly. Napr. pomocou Tukeyho procedúry.

Otázka 17. Nech $Y_{ijk} \sim N(\mu_{ij}, \sigma^2)$ je k-te pozorovanie, $k = 1, \dots, K$ na i-tej úrovni faktora A a j-tej úrovni faktora B. Zostavte zodpovedajúci model dvojfaktorovej analýzy rozptylu pre tento prípad a uveďte postup testovania hypotéz.

$$\begin{aligned} SS_A &= JK \sum_i \left(\bar{Y}_{i..} - \bar{Y}_{...}\right)^2 \\ SS_B &= IK \sum_j \left(\bar{Y}_{.j.} - \bar{Y}_{...}\right)^2 \\ SS_{AB} &= K \sum_i \sum_j \left(\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}\right)^2 \\ SS_E &= \sum_i \sum_j \sum_k \left(Y_{ijk} - \bar{Y}_{ij.}\right)^2 \end{aligned}$$

1. $H_{AB}: \ (\alpha\beta)_{ij} = \mu_{ij} - \frac{1}{J}\sum_j \mu_{ij} - \frac{1}{I}\sum_i \mu_{ij} + \frac{1}{IJ}\sum_i\sum_j \mu_{ij} = 0$
Slovne: Testujeme iterakciu i-tej úrovne A s j-tou úrovňou B.
Testovacia štatistika je

$$F_{AB} = \frac{\frac{SS_{AB}}{(I-1)(J-1)}}{\frac{SS_E}{IJ(k-1)}} \sim F\left((I-1)(J-1), IJ(k-1)\right)$$

2. Ak interakcie nie sú významené:
$H_A: \ \alpha_i = \frac{1}{J}\sum_j \mu_{ij} - \frac{1}{IJ}\sum_i\sum_j \mu_{ij} = 0$
Slovne: Testujeme efekt faktora A na i-tej úrovni.

$$F_A = \frac{\frac{SS_A}{(I-1)}}{\frac{SS_E}{IJ(k-1)}} \sim F\left((I-1), IJ(k-1)\right)$$

$H_B: \ \beta_j = \frac{1}{I}\sum_i \mu_{ij} - \frac{1}{IJ}\sum_i\sum_j \mu_{ij} = 0$
Slovne: Testujeme efekt faktora B na j-tej úrovni.

$$F_B = \frac{\frac{SS_B}{(J-1)}}{\frac{SS_E}{IJ(k-1)}} \sim F\left((J-1), IJ(k-1)\right)$$

3. Zistiť, ktoré rozdiely $\alpha_i - \alpha_j$, $\beta_i - \beta_j$ sú rôzne od nuly. Napr. pomocou Tukeyho procedúry.

Otázka 18. Nech Y_{ij} je j-te pozorovanie, $j = 1, \dots, n_i$ v i-tom základnom súbore $N_p(\mu_i, \Sigma)$, $i = 1, \dots, I$. Testujeme hypotézu $H_0: \ \mu_1 = \dots = \mu_I (= \mu)$. Zostavte zodpovedajúci model viacrozmernej jednofaktorovej analýzy rozptylu pre tento prípad a uveďte postup testovania hypotéz.

$$\begin{aligned} H &= \sum_i^k n_i \left(\bar{Y}_{.i} - \bar{Y}_{..}\right)\left(\bar{Y}_{.i} - \bar{Y}_{..}\right) \\ E &= \sum_i^k \sum_j^n \left(Y_{ij} - \bar{Y}_{.i}\right)\left(Y_{ij} - \bar{Y}_{.i}\right)^T \end{aligned}$$

1. $H_0: \ \mu_1 = \dots = \mu_I (= \mu)$
Testovacia štatistika je

$$\frac{E}{|E + H|} \sim \Lambda\left(p, n-k, n-k\right)$$

2. Zistiť, ktoré rozdiely $\mu_i - \mu_j$ sú rôzne od nuly. Napr. pomocou Tukeyho procedúry.

Otázka 19. 32 zdravých a 32 chorých ľudí sme podrobili klinickému testu pozostávajúceho zo štyroch položiek. Príslušné výsledky podľa zdravotného stavu boli nasledovné: $\bar{Y}_s = (12, 14, 17, 22)$, $\bar{Y}_v = (15, 16, 23, 27)$. Popíšte hypotézy profilovej analýzy, ktoré by sme v tomto prípade mohli testovať (predpokladajte, že nezamietame hypotézu o rovnobežnosti profilov).
Nech

$$C = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}$$

Ďalej $n_1 = n_2 = 32$ a $p = 4$.
1. Rovnobežnosť profilov: $H_0: \ C(\mu_s - \mu_v) = 0$
Testovacia štatistika je

$$\frac{n_1 n_2 (n_1 + n_2 - 2)}{(n_1 + n_2)^2} [C(\bar{Y}_s - \bar{Y}_v)]^T (CSC)^{-1} [C(\bar{Y}_s - \bar{Y}_v)] \ \sim \ T^2(p-1, n_1 + n_2 - 2)$$

2. Rovnosť dvoch úrovní: $H_0: \ 1_4^T(\mu_s - \mu_v) = 0$
Testovacia štatistika je

$$\frac{n_1 n_2 (n_1 + n_2 - 2)}{(n_1 + n_2)^2} \frac{[1_4^T(\bar{Y}_s - \bar{Y}_v)]}{1_4^T S 1_4} \ \sim \ T^2(1, n_1 + n_2 - 2)$$

3. Efekt ošetrenia: $H_0: \ \frac{1}{2} C(\mu_s + \mu_v) = 0$
Testovacia štatistika je

$$\frac{n_1 n_2 (n_1 + n_2 - 2)}{(n_1 + n_2)^2} [C\bar{Y}]^T (CSC)^{-1} [C\bar{Y}] \ \sim \ T^2(p-1, n_1 + n_2 - 2)$$

Otázka 20. Zostavte model pre situáciu, v ktorej máme súčasne kvalitatívne aj kvantitatívne premenné a popíšte postup testovania hypotéz v tomto prípade.
Model je

$$Y = X\beta + C\gamma + \varepsilon$$

kde X je matica analýzy rozptylu, C je k-ty stĺpec: n-pozorovaní k-tej kvantitatívnej premennej. Stĺpce C sú lineárne nezávislé od stĺpcov X.
Postup testovania hypotéz:
1. Testujeme vplyv sprievodných premenných na pozorovania, t. j. $H_0^1: \ \gamma = 0$
Testovacia štatistika je

$$F = \frac{n - r - t}{t} \frac{H_1 - T}{T} \sim F(t, n - r - t)$$

kde

$$\begin{aligned} h(X) &= r \\ h(C) &= t \\ H_1 &= \min_{\beta}(Y - X\beta)^T(Y - X\beta) \\ T &= \min_{\beta,\gamma}(Y - X\beta - C\gamma)^T(Y - X\beta - C\gamma) \end{aligned}$$

2. $H_0^2: \ \alpha_1 = \cdots = \alpha_p$, resp. $H_0^2: \ LB = 0$
Testovacia štatistika je

$$F = \frac{n-r-t}{s}\frac{H_2 - T}{T} \sim F(s, n-r-t)$$

kde

$$H_2 = \min_{LB=0}(Y - X\beta - C\gamma)^T(Y - X\beta - C\gamma)$$